

The invention claimed is:

- 1 1. A method of document expansion for a speech retrieval document
2 by a recognizer, comprising the steps of:
3 accessing a database of vectors of automatic transcriptions of documents;
4 truncating the vectors by removing all terms in the vectors that are not
5 recognized by the recognizer, thereby creating truncated vectors;
6 changing weights of terms in the truncated vectors to weights associated
7 with the vectors before the vectors were truncated to form the truncated vectors, thereby
8 creating truncated, weighted vectors; and
9 adding to the truncated, weighted vectors any terms which were not
10 recognized by the recognizer.
- 11 2. The method recited in claim 1, further comprising the step of
12 comparing from the truncated vectors a retrieval of documents from the original vectors,
13 thereby measuring effect of deletions from the original vectors on retrieval accuracy.
- 14 3. The method recited in claim 1, further comprising the step of
15 measuring incremental loss in retrieval effectiveness due to insertion of the terms not
16 recognized by the recognizer.
- 17 4. The method recited in claim 3, further comprising the step of
18 determining final retrieval effectiveness of the speech retrieval document using automatic
19 transcriptions.
- 20 5. The method recited in claim 4, wherein the accessing step
21 comprises the step of querying the database with the speech retrieval document to
22 retrieve documents that are similar to each other.
- 23 6. The method recited in claim 5, wherein the querying step
24 comprises retrieving a predetermined number of documents which are most similar to the
25 speech retrieval document.
- 26 7. The method recited in claim 6, wherein the predetermined number
27 is ten.
- 28 8. The method recited in claim 6, wherein the truncating step
29 comprises the step of modifying the original vectors according to a weighting function to
30 produce the truncated vectors.

1 9. The method recited in claim 8, wherein the weighting function
2 comprises a function of the following form:

3

$$\bar{D}_{new} = \alpha \bar{D}_{old} + \frac{\sum_{i=1}^{10} D_i}{10}$$

- 1 10. A system for document expansion of a speech retrieval document
2 for a recognizer, comprising:
3 a database of vectors of automatic transcriptions of documents; and
4 a processor in data communication with the database of vectors for
5 truncating the vectors by removing all terms in the vectors that are not recognized by the
6 recognizer, thereby creating truncated vectors, changing weights of terms in the truncated
7 vectors to weights associated with the vectors before the vectors were truncated to form
8 the truncated vectors, thereby creating truncated, weighted vectors, and adding to the
9 truncated, weighted vectors any terms which were not recognized by the recognizer.
- 10 11. The system recited in claim 10, wherein the processor is further
11 operable to compare from the truncated vectors a retrieval of documents from the original
12 vectors, thereby measuring effect of deletions from the original vectors on retrieval
13 accuracy.
- 14 12. The system recited in claim 10, wherein the processor is further
15 operable to measure incremental loss in retrieval effectiveness due to insertion of the
16 terms not recognized by the recognizer.
- 17 13. The system recited in claim 12, wherein the processor is further
18 operable to determine final retrieval effectiveness of the speech retrieval document using
19 automatic transcriptions.
- 20 14. The system recited in claim 13, wherein the processor is further
21 adapted retrieve documents that are similar to each other during querying.
- 22 15. The system recited in claim 14, wherein during querying of the
23 database by the processor, a predetermined number of documents are retrieved which are
24 most similar to the speech retrieval document.
- 25 16. The system recited in claim 15, wherein the predetermined number
26 is ten.
- 27 17. The method recited in claim 15, wherein during truncating the
28 original vectors are modified according to a weighting function to produce the truncated
29 vectors.
- 30